Merger of Ocean Color Data from Multiple Satellite Missions within the SIMBIOS Project

Ewa J. Kwiatkowska*, Giulietta S. Fargion**
Science Applications International Corporation, SIMBIOS Project, Code 970.2, NASA Goddard Space Flight Center

ABSTRACT

The purpose of data merger activities undertaken by the National Aeronautic and Space Administration's (NASA) Sensor Intercomparison and Merger for Biological and Interdisciplinary Studies (SIMBIOS) Project is to create scientific quality ocean color data encompassing measurements from multiple satellite missions. The fusion of data from multiple satellites will improve the quality of ocean color products over single-mission data sets by expanding spatial and temporal coverage of the world's oceans and increasing statistical confidence in generated parameters. The merger will also support a variety of new applications by taking advantage of sensor-varying calibration, spectral, spatial, temporal, and ground coverage characteristics. Leading to the data merger goals, the SIMBIOS Project has established a thorough ocean color validation program and has been cross-comparing and cross-calibrating sensor data with *in situ* measurements and data among the missions. The SIMBIOS Science Team has been studying data merger algorithms based on spectral data assimilation and spatial interpolation. The SIMBIOS Project Office has implemented statistical objective analysis and regression techniques based on artificial neural networks and support vector machines. The accuracy of the merger methods will be evaluated using *in situ* data, statistical analyses, and simple chlorophyll means – the method already implemented within the SIMBIOS Project. This paper defines challenges and suggests solutions for data merger based on the example of daily chlorophyll concentration products from Moderate Resolution Imaging Spectroradiometer (MODIS) and Sea-viewing Wide Field-of-view Sensor (SeaWiFS).

Keywords: ocean color, data merger, SeaWiFS, MODIS, SIMBIOS, neural networks, support vector machines.

1. INTRODUCTION

Phytoplankton are the principal source of organic matter in the oceans which sustain the marine food chain. They also act as a biological pump which sequesters carbon dioxide from the atmosphere to the deep ocean [1]. Some characteristics of the upper ocean, including phytoplankton concentrations, can be differentiated in terms of solar radiance scattered upward in the visible part of the electromagnetic spectrum. Phytoplankton concentrations are expressed in terms of the concentration of the main phytoplankton photosynthetic pigment, chlorophyll-a, which is often considered as an index of phytoplankton biomass [2]. Other factors which influence the backscatter signal are scattering by inorganic suspended material, scattering from water molecules, absorption by yellow substances, and reflection off the sea bottom. Remote sensing satellites can detect these spectral-radiance signatures of the ocean surface. Phytoplankton-dedicated satellite sensors operate using a range of spectral bands within the visible and near-infrared spectrum and the derived data are called ocean color. Satellite data need to undergo challenging sensor calibrations and validations [3,4], atmospheric and other corrections [5], and normalizations for satellite viewing and sun angles before water-leaving radiances can be obtained. Normalized water-leaving radiances (nLw) are then converted to chlorophyll concentrations using empirical algorithms [6].

There have been many satellite ocean color sensors on orbit and many are planned for the near future. Coastal Zone Color Scanner (CZCS) was an initial proof of concept mission. The sensor was launched by NASA in 1978 and operated for 8 years providing global coverage in ocean color data. The subsequent ocean color missions are shown in Table 1.

*ewa@simbios.gsfc.nasa.gov, phone 1 301 286 2980; ** project@simbios.gsfc.nasa.gov, phone 1 301 286 0744; http://simbios.gsfc.nasa.gov; fax 1 301 286 0268; NASA Goddard Space Flight Center, Code 970.2, Greenbelt, MD 20771-5000, USA.

Global coverage ocean color missions	Limited coverage ocean color missions
Ocean Color and Temperature Scanner (OCTS), Japan, 1996-1997, 10:41am Equator crossing local time, descending orbit	Modular Optoelectronic Scanner (MOS), Germany/India, 1996-on
Polarization and Dimensionality of the Earth's Reflectances (POLDER I), France, 1996-1997, 10:41am Equator crossing local time, descending orbit	Ocean Color Monitor (OCM), India, 1998-on
Sea-viewing Wide Field-of-view Sensor (SeaWiFS), US, 1997-on, 12:20pm Equator crossing local time, descending orbit	Ocean color Imager (OCI), Taiwan, 1999-on
Moderate Resolution Imaging Spectroradiometer AM on Terra (MODIS), US, 1999-on, 10:30am Equator crossing local time, descending orbit	Ocean Scanning Multispectral Imager (OSMI), South Korea, 1999-on
Multi-angle Imaging SpectroRadiometer (MISR), US, 1999-on, 10:30am Equator crossing local time, descending orbit	
MODIS PM on Aqua, (US) 2002-on, 1:30pm Equator crossing local time, ascending orbit	
Medium Resolution Imaging Spectrometer (MERIS), France, 2002-on, 10:00am Equator crossing local time, ascending orbit	
Global Imager (GLI), Japan, to be launched in 2002, 10:30am Equator crossing local time, descending orbit	
POLDER II (France) to be launched in 2002, 10:30am Equator crossing local time, descending orbit	
Visible/Infrared Imager Radiometer Suite (VIIRS), US, to be launched in 2006, 10:30am Equator crossing local time, descending orbit	

Table 1: Ocean color sensors.

All ocean color sensors are placed on polar orbiting satellites. Global missions revisit each worldwide location every second day while the revisits of the limited coverage missions are restricted to given imaging schedules and target locations. The global sensors cover the Earth each day with a number of consecutive swaths where the swath widths range from 1400km (OCTS) to 2400km (POLDER) and spatial resolutions per pixel vary from 250m (MODIS and GLI) to 7km (POLDER).

The ultimate objective of the NASA’s SIMBIOS Project at Goddard Space Flight Center (GSFC) is to integrate information from multiple satellite ocean color sensors and to create scientific quality data sets for routine distribution to the user community [7]. Leading to this goal, the SIMBIOS Project has been involved in work with various ocean color missions. The SIMBIOS Project Office together with the SeaWiFS Project automatically capture and store data from SeaWiFS and MOS instruments as well as download MODIS chlorophyll concentration high-level products from the MODIS Adaptive Processing Systems (MODAPS). The Project has acquired expertise with reading and analyzing data formats from different missions, including CZCS, OCTS, POLDER, MOS, OSMI, SeaWiFS, and MODIS. The Project also operates a thorough ocean color validation program to quantify the accuracies of the missions’ products in comparison to *in situ* measurements [8]. The SeaWiFS Bio-optical Archive and Storage System (SeaBASS) database of *in situ* ocean color and atmospheric observations is a compelling source of marine information for scientists around the globe [9] and will be used in the validation of data merger algorithms. The SIMBIOS Team in cooperation with the International Ocean Color Coordinating Group (IOCCG) has defined approximately 38 diagnostic ocean sites around the globe. For these sites, data are collected from different sensors to facilitate the merger activities. Current sensors collecting data within the diagnostic sites include MODIS and SeaWiFS [10]. Cross-sensor comparisons are being made to evaluate product differences between the missions, such as MODIS and SeaWiFS [11]. Software for uniform calibration and processing of selected ocean color sensors has been created to improve the level of compatibility among products. SeaWiFS nLw and aerosol optical thickness (AOT) data were used to calibrate MOS radiances [12]. OCTS and POLDER ocean color data were compared and calibrated vicariously using contemporary *in situ* measurements [13]. A comparative study and inter-calibration were preformed between OSMI and SeaWiFS [14]. The entire data set of OCTS Global Area Coverage (GAC) 4km-resolution files was reprocessed by the SeaWiFS and SIMBIOS Projects in collaboration with the National Space Development Agency of Japan (NASDA) and Japanese scientists and was made available to the user community through GSFC Distributed Active Archive Center (DAAC) [15]. CZCS, OCTS, MOS, OSMI, and SeaWiFS data are available for processing and display via SeaWiFS Data Analysis System (SeaDAS) [16]. MODIS imagery can now be displayed in SeaDAS [17]. Studies on the diurnal chlorophyll concentration variabilities are planned for the near future using data from MODIS on Terra and Aqua platforms and SeaWiFS. Finally, the Project Office has initiated research and development of methodologies for generating merged multi-sensor ocean color products. The main goal has been to define algorithms which can uniformly overcome mission-specific parameters and be applied to different products and sensors.

The most obvious benefit of the data merger is improvement in spatial and temporal ocean color coverage. Single sensor daily coverage is severely limited by gaps between the consecutive swaths and gaps caused by clouds, sun glint and other phenomena which disable the extraction of ocean color [18]. For example, merged data from three global satellite sensors MODIS on Terra and Aqua platforms and SeaWiFS provide only about 30% of the global ocean coverage at 9km resolution within a single day [19]. The other critical benefit is an increase in statistical confidence in extracted bio-optical parameters [20]. Fusion of data from multiple sensors will enable a definition of a variety of ocean color products, including long-term time series and climatological data sets. Ocean color satellite sensors are characterized by different calibration and validation accuracies and different spectral, spatial, temporal, and ground coverage attributes. Merger algorithms will take advantage of these sensor-dependent characteristics that will enable broadening the scope of ocean color applications [21].

Several merged ocean color products are expected to be produced in the near future. The design of regional and local merged ocean color products has been studied [22]. However, emphasis is currently being placed on the creation of daily global chlorophyll concentration maps. The reminder of this paper will be devoted to defining challenges and suggesting solutions for data merger based on the example of global observations from Terra's MODIS and SeaWiFS [23].

2. MODIS AND SeaWiFS DATASETS

MODIS on the Terra satellite and SeaWiFS are two global ocean color sensors currently on orbit. Their data are used to investigate merger techniques leading to improvement of daily global ocean color coverage. MODIS was launched in December 1999 and SeaWiFS in August 1997. Both sensors are in descending sun-synchronous, near-polar, circular orbits with a 10:30am local Equator crossing time for MODIS and a 12:20-noon crossing time for SeaWiFS. MODIS is equipped with a scan mirror assembly [24]. The assembly contains a continuously rotating double-sided scan mirror which scans the ground track with 10 detectors from West to East $\pm 55^0$ off nadir. The width of the MODIS swath is 2330 km cross track. The SeaWiFS assembly is composed of an off-axis folded rotating scanning telescope coupled with a half-angle scan mirror [25]. Although the maximum scan angle of SeaWiFS is $\pm 58.3^0$ at MODIS-like altitude, the SeaWiFS GAC swath is limited to 1502 km. Both sensors provide global coverage of the Earth with around 14 swaths (i.e. orbits) per day. Gaps between consecutive orbits are filled on the next day. MODIS is equipped with 36 spectral bands from which 9 are used for ocean color studies and SeaWiFS has 8 spectral bands – all of which are devoted to ocean color research. Two MODIS bands are imaged at a nominal spatial resolution of 250m at nadir, five bands at 500m, and the remaining 29 bands, including those used for ocean color, at 1km. SeaWiFS spatial resolution for all of its 8 bands is 1.1 km at nadir, however GAC data used in this study is subsampled every 4^{th} pixel and 4^{th} line giving SeaWiFS 4km resolution. MODIS standard ocean products include nLw at bands 412nm, 443nm, 488nm, 531nm, 551nm, 667nm, and 678nm; AOT at 865nm (AOT-865); diffuse attenuation coefficient at 490nm (K-490); aerosol radiance ratio between the two near-infrared bands 765nm and 865nm (ε); and three chlorophyll concentration products of which chlor_a_2 represents the OC3M algorithm which is most similar to the OC4v4 algorithm used to obtain SeaWiFS standard chlorophyll [6]. Among SeaWiFS standard products are nLw at bands 412nm, 443nm, 490nm, 510nm, 555nm, and 675nm; AOT-865; K-490; ε; and the chlor_a chlorophyll concentration product.

All results presented in this paper examine MODIS collection 4 reprocessing and SeaWiFS reprocessing 4 data. Each analysis investigates only good quality (i.e. quality 0) MODIS data and SeaWiFS level 3 standard quality data. All research on chlorophyll concentration data from both sensors uses MODIS chlor_a_2 and SeaWiFS chlor_a products.

3. PRE-MERGER DATA VALIDATIONS AND CROSS-CALIBRATIONS

There are many difficulties associated with ocean color data merger, including the merger of MODIS and SeaWiFS data. The sensors have varying designs and characteristics. They rely on different calibration approaches, processing algorithms, and means of vicarious calibration [3,4,24]. The same products can be derived using different bands for different sensors, which may cause incompatibilities. Discrepancies in sensor characteristics, calibrations, and data processing algorithms produce relationships between data products which are highly indefinite and sometimes contradictory. Data especially suspect are those contaminated by clouds, dust, other types of turbid atmosphere, in coastal waters, and mixed pixel representations. Another type of ambiguity arises from the fact that sensors are flown over the same regions at different times and a legitimate change in bio-optical conditions of the ocean cannot be reliably established and differentiated from instrument and calibration artifacts.

As much as merger activities need to depend on the calibration and validation quality of data products generated by respective science teams, differences in standard products among missions need to be evaluated [11,19]. Merger can then use the results of these evaluations by cross-calibrating the sensors and choosing the most suitable merger algorithms.

The SIMBIOS Project Office has been involved in calibration and validation efforts in support of both MODIS and SeaWiFS instruments [8,19]. The techniques used in sensor validation will be later employed to evaluate MODIS and SeaWiFS merged data products. One way to assess the accuracy of ocean color data is through comparisons, called matchups, with *in situ* measurements. The Project performs routine matchups of SeaWiFS products against *in situ* observations contained in the SeaBASS database [26]. The *in situ* matchups with MODIS data have just started and the results are preliminary and not as statistically representative as for SeaWiFS. On the logarithmic scale of chlorophyll, SeaWiFS chlorophyll matchups with *in situ* measurements result in a root-mean-square (RMS) error of 0.2386 and a linear fit slope of 1.0199 for 208 *in situ* observations which have been strictly screened for quality. MODIS chlorophyll matchups give an RMS error of 0.1709 and a linear fit slope of 0.7125 for 23 *in situ* observations. Another benchmark for merger evaluation will be provided by daily chlorophyll means of products from both sensors [19]. The means are based on binning of combined MODIS and SeaWiFS chlorophyll concentrations at 9km resolution. This joint binning produces MODIS and SeaWiFS chlorophyll averages within these 9km bins where the coverage from both instruments overlaps and a single sensor average within the bins where only data from a single instrument are present. This basic merged product will soon be available from the DAAC and through the SeaWiFS and SIMBIOS web pages.

The SIMBIOS Project has identified common data formats and products from MODIS and SeaWiFS for the comparisons, cross-calibrations, and merger. The analyses use global products where original pixel data are spatially averaged within 4.6km equal-area bins and the bins are temporally averaged within one day. Spatial and temporal binning is the same for both sensors [27] to assure exactly the same ground coverage from both instruments. Although the MODIS standard bin size is 4.6km while SeaWiFS is 9.2km, existing procedures can bin SeaWiFS data to 4.6km to facilitate the comparisons with MODIS standard products. For selected analyses daily 36km bin size data are examined and for global time series studies 9km data are used which are binned over 8 days to limit the computational effort.

The SIMBIOS Project Office has developed software for the combined extraction and analysis of binned MODIS and SeaWiFS files. The evaluation of data products from both sensors has involved:

1. Maps of chlorophyll differences between the two sensors for each day and for the 4 seasons.
2. Matchups of product data from both sensors for daily global overlaying bin coverage and for the overlaying coverage limited to open-ocean and clear-atmosphere.
3. Data product histograms.
4. Time trends in matchup statistics.
5. Separate matchups and histograms for different sensor view angles for daily global overlaying bin coverage and for the overlaying coverage limited to open-ocean and clear-atmosphere.
6. Time trends in matchup statistics for different sensor view angles.

7. Analyses of individual scan line variabilities.
8. Global time series comparisons for clear waters, deep waters, and coastal waters.

The results of these analyses have been presented to the ocean color community in a number of presentations and via the SIMBIOS Project's web sites [19,21,28,29].

Concluding, the groundwork for ocean color data merger should include thorough sensor validations and cross-comparisons which have a critical importance to the merger accomplishment. These activities enable estimation of the accuracy of sensor data involved in the merger as well as extraction of disparate trends and trend dependencies in data from different instruments. The estimates of sensor relative accuracies are then taken into consideration when integrating multi-source data. The histograms enable the examination of data distributions. The product comparisons give information on the transfer functions for data between the sensors. Distinct and quantifiable trends in instrument data will have to be corrected before integrating data. This is because propagating prominent sensor discrepancies onto merged products can cause spatial and temporal product inconsistencies. In the case of MODIS and SeaWiFS, pre-merger validations and cross-calibrations have included matchups with *in situ* measurements and matchups of products between the two sensors. The matchups of products between the sensors resulted in the extraction of
1. temporal trends,
2. and scan-angle dependencies

in the data sets which should be eliminated before the merger using sensor cross-calibrations. As an example, using statistics derived from MODIS and SeaWiFS scan line cross-calibration, MODIS data have been corrected depending on their viewing angle. The correction performed on a single day of observations improved sensor chlorophyll concentration matchups based on the slope of 0.934 in the linear fit between the two data sets to the slope of 0.984.

4. DATA MERGER APPROACHES

Merger algorithms for ocean color data should ideally meet the following requirements:

1. overcome differences in sensor characteristics, instrument calibration, and data processing algorithms,
2. assure a consistent accuracy for all merged data points,
3. incorporate product accuracy levels based on matchups with *in situ* data.

The SIMBIOS Science Team and the SIMBIOS Project Office have been investigating a number of merger methodologies. One of the applications has been the merger of satellite and *in situ* observations. The techniques have been implemented based on a semi-analytical algorithm, which produces chlorophyll concentration values from nLw data incorporated from different sources at different wavelengths [30]; a blended analysis [31], which assumes that *in situ* measurements define a 'benchmark' internal boundary condition for the spatially continuous satellite chlorophyll field [32]; and a wavelet multiresolution analysis [33], which preserves high frequency spatial variation in a satellite chlorophyll field while updating the low frequency coefficients to reflect the *in situ* values [21]. Merger opportunities for multiple ocean sensors of different spatial resolutions have been also studied in order to provide useful tools for scientists interested in local-scale geophysical phenomena. The wavelet transform is used on overlapping scenes to extract high frequency spatial chlorophyll concentration detail from a higher resolution instrument and combine this detail with the other sensor's lower resolution imagery [22]. This enables enhancement of oceanic features in lower resolution scenes through the use of higher resolution observations which is desirable when the lower resolution data are of a superior accuracy while the higher resolution data can contribute useful information about ocean color low-scale variabilities. This wavelet multiresolution merger has been studied using 1km SeaWiFS and 0.5km MOS imagery. The major emphasis of the SIMBIOS Project has been, however, placed on the creation of multi-mission daily global chlorophyll concentration maps based on global-observing sensors. This objective has been investigated using data from Terra's MODIS and SeaWiFS instruments [23].

4.1 Algorithms for merger of daily global chlorophyll concentration data

The approaches to merge daily global chlorophyll concentration products from MODIS and SeaWiFS have included:

1. Blended analysis [31]. A Poisson equation is solved over the shape-of-the-field defining chlorophyll data from one sensor given an internal boundary condition delimited by data from the other sensor, where the second sensor is assumed more valid [32].

2. Averaging. Joint binning of MODIS and SeaWiFS chlorophyll data at 9km resolution is already operational [19].
3. Weighted averaging. Weights are defined using relative accuracy levels for sensor products. These accuracy levels are derived from matchups with *in situ* measurements.
4. A semi-analytical optical algorithm [30]. The method produces chlorophyll concentrations, combined detrital particulate and dissolved absorption coefficients, and particulate backscattering coefficients from nLw data incorporated at different MODIS and SeaWiFS wavelengths for overlapping data from both sensors [34].
5. Statistical objective analysis and 3-dimensional statistical objective analysis. The method is a spatial interpolation approach.
6. Backpropagation-neural-network and support-vector-machine regression mapping. The approach emulates the response from one sensor given data from the other sensor [21].

The choice of a merger algorithm to integrate MODIS and SeaWiFS daily chlorophyll concentration is heavily dependent on the characteristics of both sensors and their diurnal coverages. MODIS and SeaWiFS are just two measurement sources with imprecise statistical properties, non-smooth chlorophyll fields, and irregular daily global distributions. The amount of joint MODIS and SeaWiFS coverage depends on the sensors' swath phase difference. The number of overlapping bins filled per day by both instruments averages around 50% of SeaWiFS and 30% of MODIS bins at 9km resolution and this common coverage becomes smaller at higher resolutions and larger at lower resolutions. Consequently, over 75% of bins at 9km resolution from the joint MODIS and SeaWiFS daily coverage does not overlap.

The blended analysis was originally designed to integrate *in situ* and satellite observations [31]. It assumes that the *in situ* data are correct point measurements and satellite data may be imprecise but that they define well the spatial variability of a given geophysical field. The method requires an extensive network of global *in situ* observations to form an internal boundary condition encircling shape-of-the-field defining satellite data. An extension of this method has been investigated which replaces *in situ* data with data obtained from a higher-quality satellite sensor and solves the Poisson equation over the chlorophyll field of a lower-quality sensor [32]. The algorithm requires well enclosing data boundaries and relatively smooth spatial geophysical fields to solve the differential equation. This poses a problem for MODIS and SeaWiFS daily chlorophyll concentration distributions which are patchy and have frequent discontinuities.

The algorithms including averaging, weighted averaging, and semi-analytical optical are well suited to perform data merger when there is overlapping coverage from both sensors. In areas observed by just a single instrument, this instrument's data are produced as a merged output. Therefore, for non-overlapping data, averaging and semi-analytical optical algorithms become redundant. MODIS and SeaWiFS pre-merger validations show that there is a lot of scatter in comparisons even when the cross-calibration is applied. Remaining differences in sensor characteristics, calibrations, and data processing algorithms produce relationships between the two data sets which are highly indefinite and sometimes random. Another type of ambiguity arises from the fact that both sensors are flown about 2 hours apart and the change in ocean bio-optical conditions cannot be reliably established and differentiated from instrument and calibration artifacts. These dissimilarities between MODIS and SeaWiFS data cause that the averaging and optical merger algorithms produce discontinuities in merged products in areas where ocean coverage by a single sensor abuts joint satellite coverages or a single coverage from the other sensor. This creates merged products of different accuracies for different data points. Some points have the accuracy of one sensor, some have the accuracy of the second sensor, and some posses a mixed accuracy.

The averaging, weighted averaging, and semi-analytical optical algorithms can be supplemented by spatial interpolation and regression mapping. The goal of the interpolation and mapping approaches is high-quality emulation of corresponding missing data to provide double-sensor coverage for all or a majority of data points. This will ensure a relatively consistent accuracy for all data points in the merged products. The method can be easily extended when merging data from more than two satellite sensors and may only become redundant if each data point has a statistically adequate number of observations from multiple instruments so that straight averaging and optical algorithms become sufficient. The reminder of this paper will concentrate on these interpolation and mapping algorithms.

4.2 Statistical objective analysis

Statistical objective analysis is derived from the topic of objective analyses which use numerical methods to estimate geophysical field variables on surfaces and on three- and four-dimensional grids from data available at discrete locations and times. The method was first applied to ground and satellite measurements in meteorology. It calculates an interpolated grid-point value as a weighted linear combination of observations [35]. In an empirical linear interpolation, weights are either a function of separation between analysis and observing locations or a function of accuracy of one observation relative to another [20]. A distance weighting with weight normalization is the most common. In the statistical objective analysis, the approximation is obtained by making additional use of the ensemble spatial correlation structure of the whole field, i.e. the spatial distribution of observations relative to one another [20]. The analysis considers instrument errors and other variations in data so that an interpolated value of a field variable does not have to be identical with an observed value at corresponding space/time coordinates, but it is intended to coincide with the signal component of the observed variable. The approach is independent of sensor-to-sensor differences in instrument design and characteristics, calibration peculiarities, and data processing. It interpolates missing sensor data in regions where only a single sensor's coverage exists so that averaging, weighted averaging, or the semi-analytical optical approach can be effectively applied on global data.

The practical requirement for the use of this algorithm is that there has to exist a preliminary "prediction" of the signal, or a first-guess field, and the objective analysis corrects this prediction by interpolating the signal with a single or multiple passes of the algorithm. At successive corrections, non-zero weights are given to observed increments only if the observations lie within a prescribed distance, known as the influence radius, of the grid point being considered. This influence radius may be decreased with successive passes of the algorithm. The analyzed grid point value is written as:

$$\boldsymbol{Y_o} = \sum_{j=1}^{m} \boldsymbol{h}_j (\boldsymbol{X}_j^o - \boldsymbol{X}_j^f) + \boldsymbol{X}_o^f \,, \tag{1}$$

where m is a number of observing locations, Y_o is the interpolated value of a grid point, X^o_j is an observation value at point j, X^f_j is a first guess for the jth observation point, X^f_o is the first guess for the analyzed grid point, and $\boldsymbol{h}$ is the weight vector. The weights are obtained by minimizing the ensemble average of the squared difference between the analysis value and the true value of the field signal [20]. The solution to this minimization problem is a covariance array for the joint distribution:

$$\begin{pmatrix} \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h_m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{11} & \cdots & \boldsymbol{\sigma}_{1\boldsymbol{m}} \\ \vdots & & \vdots \\ \boldsymbol{\sigma}_{\boldsymbol{m}1} & \cdots & \boldsymbol{\sigma_{mm}} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\sigma}_{01} \\ \vdots \\ \boldsymbol{\sigma_{om}} \end{pmatrix}, \tag{2}$$

where σ_{ij} is a covariance between an ith and jth observation.

For this study, the covariances are expressed in terms of space-lag correlation functions, ρ(s), which can be calculated from ocean color data. This assumes that the variance of the chlorophyll concentration truth-value is a constant σ^2 at all locations, the noise variance of chlorophyll is a constant η^2 and independent of location, the space-lag covariance of the chlorophyll-truth is isotropic, and the covariances of the noise at distinct locations are zero [20]. The weighting scheme for the chlorophyll-interpolated truth-value is then written as:

$$\begin{pmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \\ \vdots \\ \boldsymbol{h_m} \end{pmatrix} = \begin{pmatrix} 1+\boldsymbol{\gamma} & \boldsymbol{\rho}(\boldsymbol{s}_{12}) & \cdots & \boldsymbol{\rho}(\boldsymbol{s}_{1\boldsymbol{m}}) \\ \boldsymbol{\rho}(\boldsymbol{s}_{12}) & 1+\boldsymbol{\gamma} & \cdots & \boldsymbol{\rho}(\boldsymbol{s}_{2\boldsymbol{m}}) \\ \vdots & & \ddots & \vdots \\ \boldsymbol{\rho}(\boldsymbol{s}_{1\boldsymbol{m}}) & \boldsymbol{\rho}(\boldsymbol{s}_{2\boldsymbol{m}}) & \cdots & 1+\boldsymbol{\gamma} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\rho}(\boldsymbol{s}_{01}) \\ \boldsymbol{\rho}(\boldsymbol{s}_{02}) \\ \vdots \\ \boldsymbol{\rho}(\boldsymbol{s}_{0\boldsymbol{m}}) \end{pmatrix}, \tag{3}$$

where s_{ij} is a distance between the ith and jth points and γ^{-1} is the signal-to-noise ratio, σ^2 / η^2, of the chlorophyll concentration product. The assumptions for this equation are not met for chlorophyll concentration data. Although the interpolation is performed among data originating from the same sensor, chlorophyll truth and noise variances are difficult to establish precisely due to the modularity of ocean color calibration and processing algorithms. Chlorophyll truth and noise variances vary depending on the amount of chlorophyll concentration which fluctuates through three scales of magnitude from $0.001 mg/m^3$ to $100 mg/m^3$. The evaluation of chlorophyll accuracy comes from matchups with *in situ* measurements where the number of matchup points is below 300 for the 5-year duration of the SeaWiFS mission [26]. However, when considering chlorophyll concentration on a logarithmic scale, the range of data values shrinks substantially and chlorophyll has close to a lognormal distribution [27]. The chlorophyll and chlorophyll-error variances can then be assumed constant at all locations and the logarithmic signal-to-noise ratio can be derived from *in situ* matchup statistics.

Modeling space-lag correlation functions is a subject of extensive research [36]. The correlation is expressed as a function of the spatial separation of locations of points in geographic coordinates. For this study, the space-lag correlation function has been calculated over global daily chlorophyll concentration fields. SeaWiFS data are used for this propose at a decreased spatial resolution of 36km to limit processing time. Correlations among chlorophyll values and chlorophyll value increments at different distance ranges are obtained. A 3-dimensional statistical objective analysis is concurrently investigated using time as a third dimension and with data points separated by a day or a number of days to interpolate a given grid location. For the 3-dimensional analysis, space-lag correlations are derived for data at different distances and 1 to 7 days apart.

Figure 1 presents the distribution of the space-lag correlation functions across distances expressed in kilometers. A 15-day chlorophyll concentration mean is treated as the background first guess field. There are two functions obtained for the same day for chlorophyll concentration and for chlorophyll concentration increments from the 15-day mean. Only those bins are used in calculations which have data on at least four days out of the sequence of 15 dates. This is the correlation of increments which is used in the statistical objective analysis. The plots show mean correlations averaged over consecutive 10km distance intervals. Because of the natural spatial variability of phytoplankton concentrations, both functions do not approach the value of 1 at small distances. The spatial-lag correlations for chlorophyll increments are low, just above 0.3, and approximate 0 for distances larger than 500km. This apparent spatial diversity of chlorophyll concentration increments, averaged over global scales, makes chlorophyll data different from meteorological data, which are better correlated and for which the statistical objective analysis was originally created [20]. Shorter time series for the chlorophyll mean first guess field (e.g. 7 days) produce slightly higher correlation values but result in a less complete coverage of the global oceans to initialize the analysis. The shape of the space-lag correlation function for chlorophyll increments necessitates placing a limit on the influence radius for the analysis equal to 500km. With low correlation values and short influence radii, the statistical objective analysis may not be able to interpolate chlorophyll grid points which lie relatively far from valid data.

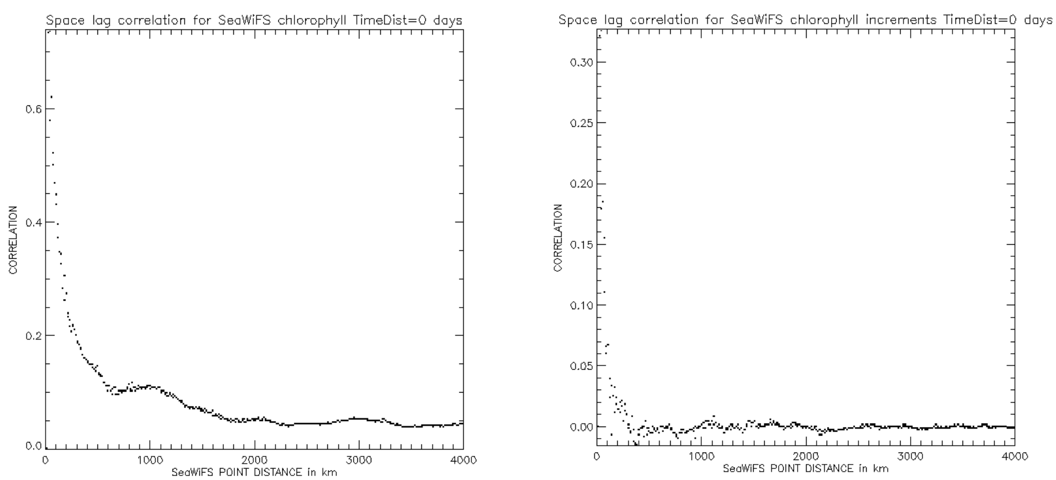

Fig. 1: Same day space-lag correlation functions averaged over 10km distances for chlorophyll concentration data and for chlorophyll increments from a 15-day mean.

The use of a mean chlorophyll field over a sequence of days is just one choice for the background first-guess fields. It may be possible to find a more suitable, better correlated, background field for the analysis. Traditionally, the first-guess would be the previous day's interpolated data. However, a spatially complete chlorophyll field for this analysis would have, again, to be someway approximated. An interesting choice would be to use the output of the neural-network or support-vector-machine regression. Otherwise, the statistical objective analysis can be modified by excluding the first guess field and interpolating the actual grid point values, not the increments from the first-guess. Then, the interpolation can still make use of ensemble spatial correlation structure of the chlorophyll field within its area of influence or completely omit the ensemble correlations and perform a straightforward linear interpolation with the space-lag statistics as weighting factors.

4.3 Neural-network and support-vector-machine regression

Because MODIS and SeaWiFS are two measurement sources with imprecise statistical properties, highly spatially variable chlorophyll fields, and patchy global distributions, an alternative merger method has been studied by the SIMBIOS Project Office. The regression approach based on neural networks and support vector machines is aimed at producing global merged ocean color products of consistent accuracy for all data points independent of their spatial locations relative to other data [21]. The essence of the approach is to reproduce the response from one sensor, given data from the other sensor. Individual data points are mapped among the sensors. The regression is performed between the first sensor's nLw at different spectral bands, chlorophyll concentration, atmospheric parameters, viewing geometry, and possible other ancillary parameters, and the second sensor's chlorophyll concentration values or, alternatively, nLw at bands used for chlorophyll calculations. Consequently, for data merger, the missing sensor's data are emulated in regions where only single sensor coverage exists so that averaging, weighted averaging, or the semi-analytical optical approach can be effectively applied on global data. The most time-consuming stages of the regression approach are the determination of an optimal set of input features, choice of the model architecture, and training on representative global and temporal data sets so as to cover the widest range of chlorophyll conditions. The algorithm learns by itself through the regression it does between both sensors' data and, at the same time, fortuitously discovers the differences in sensor data originating from sensor designs, characteristics, calibration peculiarities, and data processing. This training is highly automatic and, when once accomplished, the stored weights and system parameters can always produce new regression output with no significant processing. The method can be scaled to the case when more than two global sensors are used in the merger.

Although the mapping between sensors can be performed using linear or non-linear regression, the use of artificial neural networks or support vector machines is preferred because any complex mapping functions can be approximated using these methodologies [37]. Neural networks and support vector machines can best be understood as data transformers where the objective is to associate the elements in one set of data with the elements in the second set. They perform more accurately than other techniques, like statistical, particularly when the feature space is complex and the source data have different statistical distributions [38]. MODIS and SeaWiFS matchups show that the transfer functions between the two sensor data are convoluted and that the data are noisy and often contradictory.

Neural networks employ a connectionist approach in which complex operations are distributed among relatively simple processors, called neurons. A single neuron is stimulated by one or more inputs and it generates an output that is sent to other neurons. This output depends on the strength of each of the inputs and on the nature of the connections. There is a variety of neural networks applied today [39]. Individual neurons can be modeled by a simple weighted sum of inputs or a complex collection of differential equations. Connections between the neurons can be organized in layers such that information flows in one direction only, or it can circulate throughout the network in cyclic patterns. All neurons can be updated simultaneously, or time delays can be introduced. All responses can be deterministic or random behavior can be allowed. The neural architecture selected to perform the regression between MODIS and SeaWiFS data is a backpropagation multilayer feedforward network which uses a hyperbolic tangent as an activation function for its neurons' weighted sums of inputs. Training is accomplished using the conjugate gradient algorithm along with simulated annealing [39].

Support vector machines are learning kernel-based systems that use a hypothesis space of linear functions in high dimensional feature spaces [40]. Unlike neural networks, which try to define complex functions in the input feature space, the kernel methods perform a non-linear mapping of the complex data into high dimensional feature spaces and then use simple linear function to create linear decision boundaries. The problem of choosing an architecture for a neural network is replaced here by the problem of choosing a suitable kernel for the data projection. The advantages of support vector machines over neural networks is that they are significantly faster to train, better suited to work with high dimensional data, have only a single minimum to search, and allow for scaling the importance of outliers.

Before applying the regression technique, input data need to be prepared to simplify the training process. One important task is elimination of trends in data because the algorithm may ignore important subtle information present in the data in favor of a large variation exhibited by a trend. This is partially accomplished by MODIS and SeaWiFS cross-calibration. The other consideration is scaling. MODIS and SeaWiFS chlorophyll data are passed through a logarithm scaling function because they inherently possess a lognormal distribution [27]. All input feature data, including the chlorophyll logarithm, are then scaled and translated so that they are 0 mean and 90% of their values are within a domain range of 6.

One of the most important concerns in defining the regression model is choosing the most effective input features for the algorithm. To map MODIS data to SeaWiFS chlorophyll concentration values, dependencies between MODIS products and SeaWiFS chlorophyll have been studied using a genetic algorithm. The algorithm evaluated and propagated the fitness of various combinations of MODIS inputs through generations of neural networks trained on MODIS and SeaWiFS data sets scaled down for fast processing. The training data sets came from overlapping bins between the two sensors which were used in the matchups. MODIS input vectors were composed of nLw at all spectral bands, chlorophyll concentrations, K-490, atmospheric parameters (AOT, ε), viewing geometries (sensor and sun zenith and azimuth angles), geographic locations (latitude and longitude), and other ancillary parameters, like ozone, atmospheric pressure, humidity, wind, and date. The genetic algorithm estimated that the following MODIS input features give the best regression results in mapping to SeaWiFS chlorophyll values: nLw at bands 412nm, 488nm, 551nm, and 678nm, AOT-865, ε, chlor_a_2, satellite zenith angle, sun zenith angle, ozone, latitude, longitude, and date. Chlorophyll dependence on the viewing geometry and atmospheric and ancillary parameters means that not all of these dependencies have been eliminated in data processing and MODIS and SeaWiFS cross-calibrations. To improve the mapping, other input features such as spatial-texture and temporal chlorophyll parameters can also be considered.

5. RESULTS

The results presented here are preliminary. They are based on the application of statistical objective analysis and neural-network/support-vector-machine regression on MODIS and SeaWiFS data. The statistical objective analysis has been tested on data which are not cross-calibrated because the algorithm is not sensitive to differences between the two sensors. On the other hand, the neural-network/support-vector-machine regression is receptive to any trends in the data. Therefore, the MODIS and SeaWiFS data sets are cross-calibrated and additional features are included in the training sets such as nLw, viewing geometry, geographical location, date, and ancillary parameters as explained in section 4.3. The results presented here are obtained with MODIS and SeaWiFS daily global data binned at 36km resolution. Although, 4.6km and 9km resolution products have also been investigated, the 36km data allow for a more rapid testing of different merger application strategies and algorithm parameters.

5.1 Statistical objective analysis results

Two-dimensional and 3-dimensional forms of the statistical objective analysis have been tested as well as the versions without the background first-guess fields. Daily space-lag correlations for chlorophyll concentration data and for chlorophyll increments from a 15-day mean have been approximated using exponential functions. The function used for the interpolation of chlorophyll concentration increments within a 500km influence radius is the following:

$$Y = 0.387505 * 0.979740^{X} + 0.00785154, \tag{4}$$

and the function for the interpolation of chlorophyll without the first guess field and within a 4000km influence radius is:

$$Y = 0.759217\exp(-0.00508692 * X) + 0.0541975, \tag{5}$$

where X is a distance between points expressed in kilometers. A value of 1 is assumed when points overlap and a value of 0 is assumed when the points are outside the radius of influence. Corresponding space-lag correlation functions are obtained for points separated in time. A simple logarithmic signal-to-noise ratio of chlorophyll data is calculated by dividing the mean value of chlorophyll by the chlorophyll variance both derived from the matchups with *in situ* data.

Figure 2 shows the result of the statistical objective analysis over chlorophyll concentration increments with the background first-guess field equal to a 15-day mean and with compensation for the spatial inhomogeneity of point ensembles within radii of influence. The radius of influence is 500km. The analysis is 3-dimensional, plus and minus seven days from the examined date, 8 April 2001. SeaWiFS grid points (bins) are interpolated, however, only those bins are considered which coincide in coverage with MODIS bins containing valid data. As a similar interpolation is done on the MODIS data set, the MODIS and SeaWiFS chlorophyll concentration products can be merged using averaging or weighted averaging approaches. For the analysis to be applied with the semi-analytical optical algorithm, the interpolation has to be done on MODIS and SeaWiFS nLw at bands used for chlorophyll extraction.

Statistical objective analyses have been performed with and without the first-guess background fields, with corresponding modifications to the radii of influence. For the interpolation, problematic areas in MODIS and SeaWiFS daily imagery are those where gaps in global coverage are large, such as between sensor consecutive orbits. It has been shown that within these gaps the interpolation without the background field, although with longer influence radii, overly smoothes the chlorophyll field and can be considered impractical. The result of the conventional analysis with the first-guess field is realistic, although some interpolated coverage has values virtually repeated from the first guess field.

The statistical objective analysis is computationally involved. It considers ensemble spatial distributions of observations relative to one another which are contained within a radius of influence of an investigated grid point. This results in large covariance matrices whose size depends on the number of valid data bins within the radius of the interpolated point. At 36km resolution, there are 86782 bins of data within a 4000km radius from a central grid point. The quantity of bins is an order of magnitude higher at the resolution of 9km; where 9km is the ultimate resolution for the data merger. In practice, the number of valid data bins used in the interpolation is lower then this because it depends on the neighboring land and cloud coverage and other conditions. Nevertheless, the analysis involves inverting the square matrices of covariances for each investigated grid point. To ease the computational effort for this study, radii smaller than the influence radius are first searched to determine whether they contain a relatively high proportion of valid data bins to perform the interpolation. Because the covariance matrix is symmetric and positive definite, the Cholesky decomposition is used to solve Equation 3 [41].

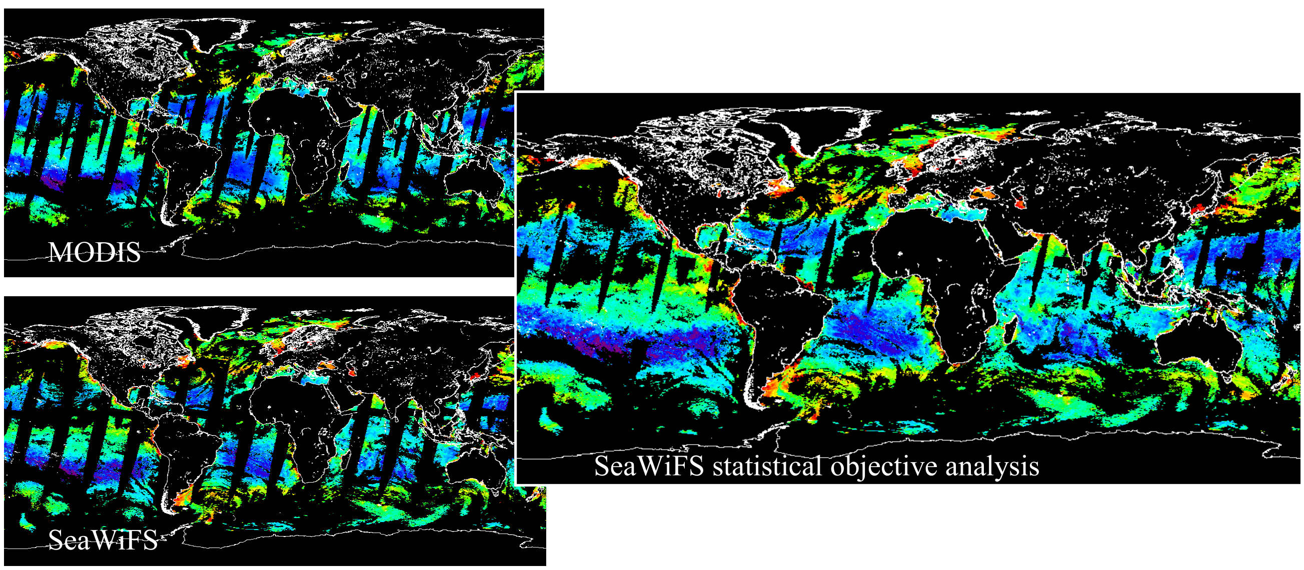

Fig. 2: Original MODIS and SeaWiFS 36km binned chlorophyll concentration data sets for 8 April 2001 and the result of the 3-dimensional statistical objective analysis on the SeaWiFS chlorophyll bins coinciding with the MODIS coverage.

More research is needed to make the statistical objective analysis more effective in terms of the choice of a chlorophyll background field and associated space-lag correlation function and influence radius, as well as the computational efficiency of the method. The statistical objective analysis does not provide inherent tools for validation of interpolation accuracy. One method of validation is to perform the interpolation on data grid points with known values. However, matchups with *in situ* measurements coupled with statistical analyses and visual inspections, which check for noticeable discontinuities in the chlorophyll fields, create more comprehensive verifications of the analysis.

5.2 Neural-network and support-vector-machine regression results

Both multilayer feedforward neural networks and support vector machines have been used to perform the regression between MODIS and SeaWiFS data. Mapping has been initially studied between MODIS nLw at bands 412nm, 488nm, 551nm, and 678nm, AOT-865, ε, chlor_a_2, satellite zenith angle, sun zenith angle, ozone, latitude, longitude, and date, and the SeaWiFS chlor_a product. The different MODIS products are used because they were determined to be correlated with SeaWiFS chlorophyll data and in turn contribute useful information for the solution of this very complex regression problem (see Sec. 4.3). Data used for the regression come from overlapping bin coverage between the two sensors. Input features are scaled and translated into 0 mean values. Studies have been performed to obtain MODIS data mapping to SeaWiFS chlorophyll running across 2 years of concurrent sensor coverage. There are some temporal trends in the data and scan angle dependencies that are still being studied. These dependencies are attributed to the change in bio-optical ocean properties revealed in the two sensors' daily observations, which are two hours apart, and also to MODIS calibration epochs. The trends limit the ability of the algorithm to descend into a stable solution and require large training data sets and long training times. These are currently being investigated on a Cray supercomputer at the GSFC NASA Center for Computational Sciences. The regression has been accomplished for the most ideal MODIS data conditions such as West-most scan edge data. The research into cross-seasonal mapping continues. The results presented here are based on the regression performed on a single-day data set. Half of the overlapping MODIS and SeaWiFS bins are used as a training set and the other half as a testing set.

The neural network topology is three layered. The number of neurons in input layer is equal to the number of MODIS input features. There is a single hidden layer. The output layer is made up of a single neuron, which is the resultant SeaWiFS chlorophyll value. A range of sizes of the hidden layer has been investigated as well as a variety of algorithm parameters. Neural networks have also been employed by the genetic algorithm to find the set of most effective input features for the regression as described in Sec. 4.3.

The convergence of support vector machines depends heavily on the choice of a kernel function. This study has found the radial basis function to be the best kernel to perform the regression between MODIS and SeaWiFS data [42]. A search has been performed for the most effective capacity parameter to define the best separation between MODIS and SeaWiFS data in kernel-projected multi-dimensional feature spaces and the most appropriate algorithm sensitivity to outliers [43].

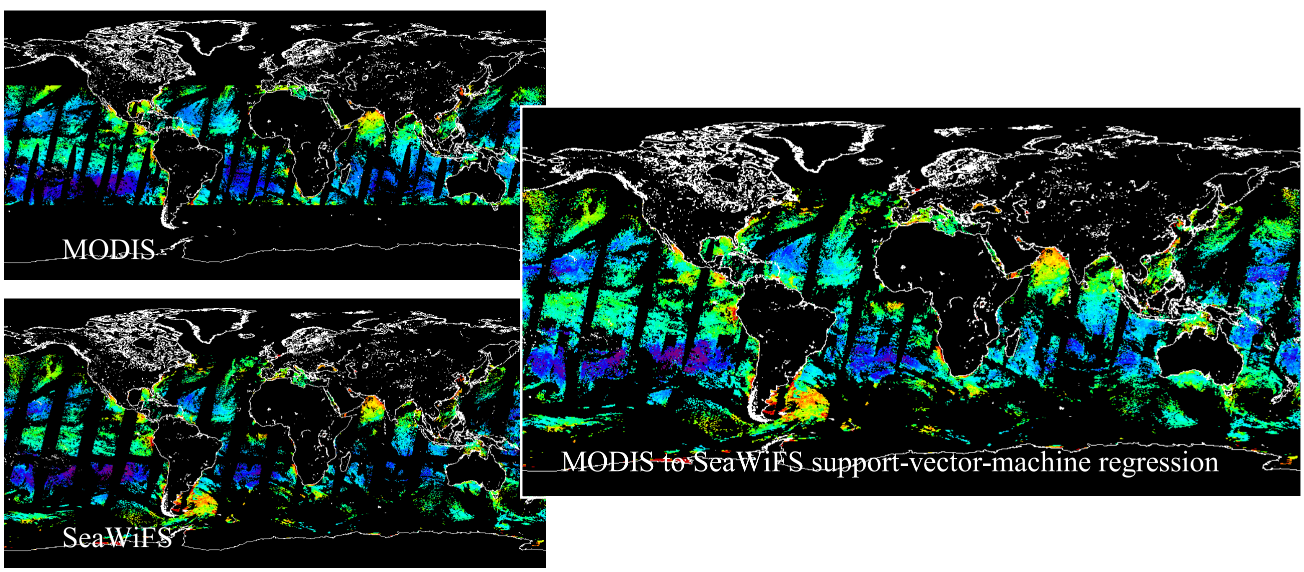

Fig. 3: Original MODIS and SeaWiFS 36km binned chlorophyll concentration data sets for 9 January 2002 and the result of support-vector-machine regression: SeaWiFS coverage supplemented by MODIS coverage mapped to SeaWiFS-like chlorophyll.

Figure 3 shows the result of support-vector-machine regression between MODIS data and SeaWiFS chlorophyll. Both sensors' daily global coverage is expressed in terms of SeaWiFS chlorophyll concentration where all coverage for this day unique to MODIS is mapped to SeaWiFS chlorophyll values. Only those MODIS bins are mapped which are within 40^0 South to 40^0 North coverage, where there is a single daily coverage and thus a unique sensor viewing geometry within the bins, and which contain required MODIS features: nLw at bands 412nm, 488nm, 551nm, and 678nm, AOT-865, ε, chlor_a_2, satellite zenith angle, sun zenith angle, and ozone. The mapping is performed on a single day of data, 9 January 2002. The accuracy of the mapping has been estimated on the testing set and its mean absolute error is equal to

0.09mg/m^3 of chlorophyll concentration. While a similar mapping is done from SeaWiFS to MODIS data, the MODIS and SeaWiFS chlorophyll concentration products can be successfully merged using averaging and weighted averaging approaches. For the regression to be applied with the semi-analytical optical algorithm, the mapping has to be to MODIS and SeaWiFS nLw values at bands used for chlorophyll calculations.

Support vector machine regression is a promising tool for data merger. Both neural networks and support vector machines provide intrinsic means for the evaluation of the accuracy of their learning. They are trained to converge to a stable solution and a given allowable error on the training set. Their knowledge is stored in system weights and can be tested anytime against data for which the result is known in advance. For the ultimate ocean color data merger, this built-in algorithm testing needs to be performed concurrently with the merged product validations through matchups with *in situ* measurements, statistical analyses, and visual inspections checking for noticeable discontinuities in the resultant product fields.

6. CONCLUSIONS

Because of the increasing number of ocean color missions launched and planned by the US and international community, data merger activities have become one of the priorities of the SIMBIOS Project [19,7]. The Project's Science Team and the Office have gained expertise in a variety of different approaches to ocean color data merger. The SIMBIOS Project has been investigating ocean color merger opportunities at local spatial scales to provide useful tools for the science community [32,30,21]. A wavelet-based algorithm to enhance oceanic features in lower resolution imagery through the use of higher resolution data has been implemented as well as merger of satellite and *in situ* measurements [22]. However, the most pressing assignment has become the production of daily global chlorophyll data sets encompassing multi-sensor measurements [23].

The number of global sensors is still not sufficient to cover the majority of the globe with a statistically viable numbers of observations for simple averaging or interpolation mergers. This paper shows that in order to make merged products more useful than any single-sensor data sets, thorough validations and cross-calibrations first need to be performed on multi-sensor data [19]. These cross-calibrations help to eliminate obvious data trends and make data sets comparable. This does not account for variabilities in bio-optical ocean properties occurring over few-hour-overflight differences between the sensors. However, as the initial merger goal is to increase global ocean color coverage, diurnal effects will be averaged out anyway. To investigate few-hour changes in ocean biology over global scales, there has to be high confidence in instrument characterization and data calibrations.

There has been a variety of methods proposed to perform the daily global merger. These methods include blended analysis, averaging and weighted averaging, semi-analytical optical algorithms, statistical objective analysis, and neural-network and support-vector-machine regression [32,34,21]. It has been shown that averaging, weighted averaging, and semi-analytical optical algorithms can be enhanced by statistical objective analysis' interpolation and regression mapping which would limit the discontinuities in resultant merged product fields. The SIMBIOS Project Office has implemented and investigated different approaches to spatial objective analysis and neural-network and support-vector-machine regression. The initial results show that these two methods can make useful tools for ocean color data merger. The statistical objective analysis is more computationally involved for operational processing of multi-sensor data because of the lengthy interpolation which needs to be performed on every merged image. It also requires more research into whether it can accurately interpolate ocean color data for large spatial gaps in coverage. The regression approaches are more computationally efficient for operational use because their classification is performed using stored weights which work like a look-up table and carry the knowledge of the mapping function. This knowledge is learned just once, although through a time-consuming and comprehensive training effort. The method is independent of spatial separation of the ocean color data points. Support vector machines with kernels composed of radial basis functions are currently found to be the most efficient in performing the regression. The results of spatial objective analysis and regression mapping have been shown using MODIS and SeaWiFS 36km daily global data sets.

The SIMBIOS Project Office will carry on its investigation into ocean color merger methodologies and will continue cooperation with the SIMBIOS Science Team, MODIS Team, and the IOCCG on sensor intercomparisons, data merger, and algorithm implementation. The developed scientific tools and expertise will be used with subsequent ocean color missions, such as MODIS-Aqua, MERIS, GLI, POLDER II, and VIIRS.

ACKNOWLEDGMENTS

This research has been supported by the NASA GSFC SIMBIOS Project. The authors would like to thank MODIS and SeaWiFS teams for their stimulating input and support and, in particular, Janet W. Campbell, Robert L. Evans, Katherine Kilpatrick, Wayne E. Esaias, Watson W. Gregg, Stéphane Maritorena, David A. Siegel, Bryan A. Franz, Sean W. Bailey, Charles R. McClain, and the GSFC DAAC team.

REFERENCES

1. M.J. Behrenfeld, J.T. Randerson, C.R. McClain, G.C. Feldman, S.O. Los, C.J. Tucker, P.G. Falkowski, C. B. Field, R. Frouin, W.E. Esaias, D.D. Kolber, and N.H. Pollack, "Biospheric Primary Production During an ENSO Transition," *Science*, **291**, pp. 2594-2597, 2001.
2. S. Sathyendranath, Ed., *Remote Sensing of Ocean Colour in Coastal, and Other Optically-Complex, Waters*, Reports of the International Ocean-Colour Coordinating Group, **3**, pp. 5-21, 2000.
3. R.A. Barnes, R.E. Eplee Jr., G.M. Schmidt, F.S. Patt, and C.R. McClain, "Calibration of SeaWiFS. I. Direct Techniques," *Applied Optics*, vol. 40, no. 36, pp. 6682-6700, 2001.
4. R.E. Eplee Jr., W.D. Robinson, S.W. Bailey, D.K. Clark, P.J. Werdell, M. Wang, R.A Barnes, and C.R. McClain, "Calibration of SeaWiFS. II. Vicarious Techniques," *Applied Optics*, vol. 40, no. 36, pp. 6701-6718, 2001.
5. H.R. Gordon and M. Wang, "Retrieval of Water-Leaving Radiance and Aerosol Optical Thickness over the Oceans with SeaWiFS: a Preliminary Algorithm," *Applied Optics*, **33**, pp. 443-452, 1994.
6. J.E. O'Reilly, S. Maritorena, B.G. Mitchell, D.A. Siegel, K.L. Carder, S.A. Garver, M. Kahru, and C.R. McClain, "Ocean Color Chlorophyll Algorithms for SeaWiFS," *J. Geoph. Research*, vol. 103, no. C11, pp. 24,937-24,953, 1998.
7. G.S. Fargion and C.R. McClain, *SIMBIOS Project 2001 Annual Report*, NASA Technical Memorandum, no. 2002-21005, pp. 1-308, 2002.
8. J.L. Mueller and G.S. Fargion, *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 3*, NASA Technical Memorandum, no. 2002-21004, vol. 1 & 2, pp. 1-184, 2002.
9. J.P. Werdell and S.W. Bailey, *The SeaWiFS Bio-optical Archive and Storage System (SeaBASS): Current Architecture and Implementation*, NASA Technical Memorandum, no. 2002-211617, G.S. Fargion and C.R. McClain Eds., 2002.
10. SIMBIOS and SeaWiFS Project web site: http://seawifs.gsfc.nasa.gov/cgi/seawifs_region_extracts.pl, 2002.
11. K. Kilpatrick, E. Kearns, E.J. Kwiatkowska-Ainsworth, and R.L. Evans, "Time Series of Calibrated Ocean Products from NASA's Moderate Resolution Scanning Spectrometer (MODIS)," *Proceedings of the Ocean Sciences Meeting*, Honolulu, Hawaii, USA, Feb. 2002.
12. M. Wang and B.A. Franz, "Comparing the Ocean Color Measurements between MOS and SeaWiFS: A Vicarious Intercalibration Approach for MOS," *IEEE Trans. Geosci. and Remote Sens*., vol. 38, no. 1, pp. 184-197, 2000.
13. M. Wang, A. Isaacman, B.A. Franz, and C.R. McClain, "Ocean-Color Optical Property Data Derivation from the Japanese Ocean Color and Temperature Scanner and the French Polarization and Directionality of the Earth's Reflectances: a Comparison Study," *Applied Optics*, vol. 41, no. 6, pp. 974-990, 2002.
14. B.A. Franz and Y. Kim, "A Comparative Study and Intercalibration Between OSMI and SeaWiFS," *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, vol. 82, no. 47, pp. F661 (OS42D-09), Dec. 2001.
15. T. Tanaka, B.A. Franz, J.G. Acker, I. Asanuma, S. Bailey, R.E. Eplee Jr. H. Fukushima, J.M. Gales, S. Maritorena, Y.Mitomi, H. Murakami, J.E. O'Reilly, S. Shen, P. Smith, M. Wang, J. Wilding, and B. Woodford, "Reprocessing of the OCTS Global Dataset, a colaborative Effort Between NASDA and the NASA SIMBIOS Project," *Eos Trans. American Geophysical Union*, vol. 82, no. 47, Fall Meeting, San Francisco, USA, pp. F660-F661, Dec. 2001.
16. G. Fu, K.S. Baith, and C.R. McClain, "SeaDAS: The SeaWiFS Data Analysis System," *Proceedings of the 4th Pacific Ocean Remote Sensing Conference*, pp. 73-79, Qingdao, China, Jul. 1998.
17. K.S. Baith, R. Lindsay, G. Fu, and C.R. McClain, "Data Analysis System Developed for Ocean Color Satellite Sensors," *Eos Trans. American Geophysical Union*, vol. 82, no. 18, 2001.
18. W.W. Gregg, W.E. Esaias, G.C. Feldman, R. Frouin, S.B. Hooker, C.R. McClain, R.H. Woodward, "Coverage Opportunities for Global Ocean Color in a Multimission Era," *IEEE Trans. Geosci. and Remote Sens*., vol. 36, no. 5, pp. 1620-1627, 1998.
19. G.S. Fargion, E.J. Kwiatkowska, and S. Bailey, "An Overview of SIMBIOS Project Contributions in Support of MODIS," *MODIS Science Team Meeting presentation*, Greenbelt, MD, USA, July 2002, http://modis-ocean.gsfc.nasa.gov/refs.html.
20. H.J. Thiebaux and M.A. Pedder, *Spatial Objective Analysis with Applications in Atmospheric Sciences*, Academic Press, pp. 129-204, 1987.

21. E.J. Kwiatkowska and G.S. Fargion, “Merger of Ocean Color Information from Multiple Satellite Missions under the NASA SIMBIOS Project Office,” *Proceedings of the Fifth International Conference on Information Fusion*, July 2002, Annapolis, MD, USA, **1**, pp. 291-298.
22. E.J. Kwiatkowska-Ainsworth, “Merger of Ocean Color Information of Different Spatial Resolution: SeaWiFS and MOS,” *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, vol. 82, no. 47, pp. F675 (OS52A-0514), Dec. 2001.
23. W.W. Gregg and R.H. Woodward, “Improvements in coverage Frequency of Ocean Color: combining Data from SeaWiFS and MODIS,” *IEEE Trans. Geosci. and Remote Sens*., vol. 36, no. 4, pp. 1350-1353, 1998.
24. W.E. Esaias, M.R. Abbot, I. Barton, O.B. Brown, J.W. Campbell, K.L. Carder, D.K. Clark, R.L. Evans, F.E. Hoge, H.R. Gordon, W.P. Balch, R. Letelier, and P.J. Minnett, “An Overview of MODIS Capabilities for Ocean Science Observations,” *IEEE Trans. Geosci. and Remote Sens*., vol. 36, no. 4, pp. 1250-1265, 1998.
25. SeaWiFS Project web site: http://seawifs.gsfc.nasa.gov/SEAWIFS/SEASTAR/SPACECRAFT.html, 2002.
26. S.W. Bailey, J.P. Werdell, and C.R. McClain, “Validation of Satellite-Derived Ocean Color: Theory and Practice,” *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, vol. 82, no. 47, pp. F676 (OS52A-0520), Dec. 2001.
27. J.W. Campbell, J.M. Blaisdell, and M. Darzi, *Level-3 SeaWiFS Data Products: Spatial and Temporal Binning Algorithms*, NASA Technical Memorandum, no. 104566, vol. 32, 1995.
28. B.A. Franz, “MODIS and SeaWiFS Global Time Series Analysis,” web site: http://simbios.gsfc.nasa.gov/~franz/l3trend/modis_seawifs/modis_seawifs_8day_trends.html, 2002.
29. E.J. Kwiatkowska, “MODIS/SeaWiFS matchups,” web site: http://simbios.gsfc.nasa.gov/~ewa/SeaMODISMay/seamodis-match.html, 2002.
30. S. Maritorena, D.A. Siegel, and A. Peterson, “Optimization of a Semi-Analytical Ocean Color Model for Global Scale Applications,” *Applied Optics*, vol. 41, no. 15, pp. 2705-2714, 2002.
31. R.W. Reynolds, “A Real-Time Global Sea Surface Temperature Analysis,” *Climate*, **1**, pp. 75-86, 1988.
32. W.W. Gregg and M.E. Conkright, “Global Seasonal Climatologies of Ocean Chlorophyll: Blending In Situ and Satellite Data for the CZCS Era,” *J. Geoph. Research – Oceans*, **106(C2)**, pp. 2499-2515, 2001.
33. S.G. Mallat, “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, 1989.
34. S. Maritorena, D.A. Siegel, A.R. Peterson and M. Lorenzi-Kayser, “Tuning of a Pseudo-analytical Ocean Color Algorithm for Studies at Global Scales,” *Eos Trans. American Geophysical Union*, OS12M-05, San Antonio, USA, Jan. 2000.
35. H.J. Thiebaux, “Maximally Stable Estimation of Meteorological Parameters at Grid Points,” *J. Atmospheric Science*, vol. 30, pp. 1710-1714, 1973.
36. P.R. Julian and H.J. Thiebaux, “On some Properties of Correlation Functions Used in Optimum Interpolation Schemes,” *Monthly Weather Review*, **103**, pp. 605-616, 1975.
37. P.M. Atkinson and A.R.L. Tatnall, “Neural Networks in Remote Sensing,” *Int. J. Remote Sens*., vol. 18, no. 4, pp. 699-709, 1997.
38. Y.-H. Pao, *Adaptive Pattern Recognition and Neural Networks*, Reading, MA, Addison-Wesley, 1989.
39. T. Masters, *Practical Neural Network Recipes in C++*, Academic Press, Inc., 1993.
40. N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.
41. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*, Cambridge University Press, 1992.
42. B. Schölkopf,, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, V. Vapnik, “Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers,” *Massachusetts Institute of Technology Artificial Intelligence Laboratory Memorandum*, no. 1599, C.B.C.L. paper no. 142, 1996.
43. B. Schölkopf, “Statistical Learning and Kernel Methods,” *Microsoft Research Technical Report*, no. MSR-TR-2000-23, pp. 1-27, 2000.